Supporting Information

Prediction of radionuclide diffusion enabled by missing data imputation and ensemble machine learning

Jun-Lei Tian[1], Jia-Xing Feng[1], Jia-Cong Shen[1], Lei Yao[1], Jing-Yan Wang[1], Tao Wu[1,*], and Yao-Lin zhao[2,*]

[1] *Huzhou Key Laboratory of Environmental Functional Materials and Pollution Control, Huzhou University, Huzhou 313000 China*
[2] *School of Nuclear Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China*

*Corresponding author：Tao Wu and Yao-Lin Zhao

Email: twu@zjhu.edu.cn (Tao Wu),
zhaoyaolin@mail.xjtu.edu.cn (Yao-Lin Zhao)

Phone numbers: 13655729651

Postal address: Huzhou University, No. 759 East Erhuan Road, Wuxing District, Huzhou, Zhejiang Province, China, 313000

*Corresponding author. EMail address: twu@zjhu.edu.cn (Tao Wu), zhaoyaolin@mail.xjtu.edu.cn (Yao-Lin Zhao)

Physical parameters

Physical parameters related to the diffusion process could help us to understand the diffusion mechanism. In this study, five physical parameters, including the total porosity, ion molar conductivity, ionic radius, and montmorillonite stacking number, were incorporated into datasets to enhance the predictive performance and provide deeper insights into the diffusion mechanism. Specifically, the total porosity (ε_{tot}) is related to the grain density of bentonite (ρ_{d}) and compacted dry density (ρ_{s}). It defines as:

$$\varepsilon_{\text{tot}} = 1 - \frac{\rho_{\text{d}}}{\rho_{\text{s}}}, \tag{S1}$$

The relationship between the rock capacity factor (α) and porosity (ε) is expressed by:

$$\alpha = \varepsilon + K_{\text{d}} \cdot \rho_{\text{d}}, \tag{S2}$$

The anionic exclusion effect has been recognized by numerous diffusion models [1, 2, 3]. It assumes that radionuclide anions cannot enter the interlayer pores of compacted bentonite due to the anionic exclusion from the negative charge of the montmorillonite layer. Consequently, the rock capacity factor of radionuclide anions should be less than the total porosity. However, the calculation of rock capacity factor in JAEA-DDB neglects the anionic exclusion effect, leading to an overestimation of this parameter for radionuclide anions. In this study, the anionic exclusion effect was considered. That is, ε=ε_{tot} in Eq. (S1) was employed for radionuclides with $K_{\text{d}} > 0$, whereas ε= 1–0.39·ρ_{d} was used for these with $K_{\text{d}} = 0$ [4].

The ion molar conductivity (λ) is expressed by the following equation [5]:

$$\lambda = \frac{D_{\text{w}} \cdot F^2 \cdot |z|}{R \cdot T}, \tag{S3}$$

where F is the Faraday constant (96500 C/mol), z is the ion charge, T is the Kelvin temperature (K), and R is the molar gas constant (8.314 J/mol·k).

Micro-parameters, such as the ionic radius and montmorillonite stacking number, were integrated into the datasets to establish a connection between mesoscopic and microscopic features. The ionic radius (r) is calculated using the classic Stokes-Einstein relation [6]:

$$r = \frac{k \cdot T}{6 \cdot \pi \cdot \eta \cdot D_{\text{w}}}, \tag{S4}$$

where k is the Boltzmann constant (1.380649×10^{23} Pa·m^3/K), η is the viscosity of the medium.

The montmorillonite stacking number is approximately given as [7]:

$$n_c = \frac{2 \cdot a \cdot b \cdot n_a \cdot n_b \cdot N_A}{MW \cdot A_{\text{ext}} \cdot n_a \cdot n_b - 2 \cdot a \cdot c^* \cdot n_a \cdot N_A - 2 \cdot b \cdot c^* \cdot n_b \cdot N_A}, \tag{S5}$$

where N_A is Avogadro's number (6.022×10^{23} molecules/mol). The chemical molecular weight (MW) of a typical montmorillonite is 735 g/mol. The unit-cell dimension of a montmorillonite layer is $a \times b \times c^* = 0.523 \times 0.905 \times 0.96$ nm^3. The length and width of a montmorillonite layer n_a and n_b were approximately 200.

Table S1 The statistical information of the Dataset I. Dataset = 316 instances.

Parameters	Mean	Min	Max	Std	Skw
Ionic strength (I, mol/L)	0.25	0.00	1.03	0.20	0.86
Compacted dry density (ρ_d, kg/m^3)	1336	400	2100	350	−0.47
Montmorillonite content, (m, −)	0.75	0.33	1.00	0.16	−1.04
Distribution coefficient, (K_d, m^3/kg)	0.06	0.00	2.12	0.20	6.71
Ionic charge (z, −)	−0.45	−2.00	5.00	1.23	1.76
Temperature (T, °C)	26.85	10.00	80.00	14.22	2.64
Specific surface area (A_{ext}, m^2/g)	46.00	25.60	99.80	15.58	0.48
Grain density (ρ_s, kg/m^3)	2745	2660	2900	50	0.03
Molecular weight (MW, g/mol)	152.94	22.00	432.34	80.37	0.78
Species diffusion coeffcient in water (LogD_w, −)	−8.79	−9.30	−8.34	0.16	−0.77
Total porosity (ε_{tot}, −)	0.51	0.24	0.86	0.13	0.49
Rock capacity factor (α, −)	76.83	0.30	4240.28	345.00	9.70
Ionic radius (Log$r$, −)	−9.84	−10.24	−9.34	0.16	0.97
Ion molar conductivity (λ, m^2·S/ mol)	0.01	0.00	0.03	0.00	3.84
Montmorillonite stacking number (n_c, −)	28	9	54	15	0.68
pH (−)	7.08	3.00	13.40	1.59	0.23

Std = Standard Deviation; Skw = Skewness

Table S2 The statistical information of the Dataset III. Dataset = 813 instances.

Parameters	Mean	Min	Max	Std	Skw
Ionic strength (I, mol/L)	0.18	0.00	1.36	0.26	2.01
Compacted dry density (ρ_d, kg/m^3)	1303	400	2300	374	0.06
Montmorillonite content, (m, −)	0.87	0.33	1.00	0.15	−1.53
Distribution coefficient, (K_d, m^3/kg)	0.03	0.00	1.34	0.13	6.13
Ionic charge (z, −)	0.07	−2.00	5.00	1.20	0.43
Temperature (T, °C)	28.27	5.00	90.00	15.66	1.36
Specific surface area (A_{ext}, m^2/g)	37.50	25.60	99.80	15.00	1.89
Grain density (ρ_s, kg/m^3)	2814	2660	2900	73	−0.53
Molecular weight (MW, g/mol)	90.26	22.00	432.34	79.47	1.37
Species diffusion coeffcient in water (LogD_w, −)	−8.79	−9.30	−8.24	0.20	−0.21
Total porosity (ε_{tot}, −)	0.54	0.16	0.86	0.14	−0.06
Rock capacity factor (α, −)	44.84	0.30	1854.35	201.59	6.79
Ionic radius (Log$r$, −)	−9.84	−10.31	−9.33	0.19	0.33
Ion molar conductivity (λ, m^2·S/ mol)	0.01	0.00	0.02	0.00	1.25
Montmorillonite stacking number (n_c, −)	37	9	54	13	−0.69
pH (−)	6.90	3.00	13.40	1.24	0.45
Effective diffusion coefficient (LogD_e, −)	−10.05	−12.57	−8.54	0.53	−0.82

Std = Standard Deviation; Skw = Skewness

Table S3 Hyperparameters, tuning range of the machine learning models, and the optimized hyperparameters for dataset I.

Algorithms	Parameter name	Range	Dataset I
LGBM	Max_depth	1−30	−1
	Learning_rate	0.01−0.1	0.01
	Min_data_in_leaf	1−30	3
	Feature_fraction	0.1−1.0	0.4
	Bagging_freq	1−100	30
	Bagging_seed	1−100	1
	Bagging_fraction	0.01−1.0	0.75
	Lambda_l1	0.0−1.0	0.01
	Lambda_l2	0.0−1.0	0.03
CatB	Depth	1−16	5
	Learning_rate	0.01−0.5	0.05
	Subsample	0.01−1	0.7
	L2_leaf_reg	0.01−1	0.03
	Rsm	0.01−1	0.88
	Random_seed	1−100	6
XGB	Max_depth	1−15	4
	Eta	0.01−0.5	0.05
	Gamma	0.01−1	0.1
	Lambda	0.01−1	0.19
	Subsample	0.01−1	0.7
	Reg_alpha	0.01−1	0.1
	Min_child_weight	1−100	13
RF	N_estimators	1−100	64
	Max_depth	1−15	9
	Min_samples_split	2−50	4
	Min_samples_leaf	1−50	1
	Min_weight_fraction_leaf	0.01−0.5	0.01
	Random_state	1−100	64

Table S4 The optimized hyperparameters used in machine learning models for dataset II and dataset III.

Algorithms	Parameter name	Dataset II	Dataset III	Parameter name	Dataset II	Dataset III
(1) LGBM-CatB	Max_depth	−1	2	Depth	4	4
	Learning_rate	0.05	0.1	Learning_rate	0.01	0.05
	Num_leaves	30	30	Iterations	2000	2000
	Min_data_in_leaf	20	21	Subsample	0.8	0.3
	Feature_fraction	0.3	0.1	L2_leaf_reg	0.01	0.01
	Bagging_freq	40	45	Rsm	0.45	0.3
	Bagging_seed	4	4	Random_seed	15	1
	Bagging_fraction	0.4	0.96			
	Lambda_l1	0.01	0.01			
	Lambda_l2	0.05	0.05			
(2) LGBM-XGB	Max_depth	−1	7	Max_depth	3	4
	Learning_rate	0.05	0.1	Eta	0.05	0.05
	Num_leaves	30	30	N_estimators	1000	1000
	Min_data_in_leaf	20	8	Gamma	0.01	0.01
	Feature_fraction	0.2	0.2	Lambda	0.09	0.09
	Bagging_freq	30	27	Subsample	0.51	0.92
	Bagging_seed	28	37	Reg_alpha	0.04	0.04
	Bagging_fraction	0.43	0.9	Min_child_weight	6	3
	Lambda_l1	0.03	0.02	Colsample_bytree	0.5	0.85
	Lambda_l2	0.05	0.65			

Table S4-Continue

Algorithms	Parameter name	Dataset II	Dataset III	Parameter name	Dataset II	Dataset III
(3) LGBM-RF	Max_depth	4	7	N_estimators	64	100
	Learning_rate	0.05	0.1	Max_depth	5	9
	Num_leaves	30	30	Min_samples_split	5	5
	Min_data_in_leaf	9	4	Min_samples_leaf	4	4
	Feature_fraction	0.3	0.1	Min_weight_fraction_leaf	0.01	0.01
	Bagging_freq	21	21	Random_state	42	49
	Bagging_seed	15	5			
	Bagging_fraction	0.5	0.43			
	Lambda_l1	0.02	0.06			
	Lambda_l2	0.08	0.03			
(4) LGBM	Max_depth	−2	4			
	Learning_rate	0.05	0.05			
	Num_leaves	30	30			
	Min_data_in_leaf	4	8			
	Feature_fraction	0.45	0.25			
	Bagging_freq	3	10			
	Bagging_seed	5	13			
	Bagging_fraction	0.90	0.90			
	Lambda_l1	0.06	0.06			
	Lambda_l2	0.03	0.08			

Table S4-Continue

Algorithms	Parameter name	Dataset II	Dataset III
(5) CatB	Depth	2	4
	Learning_rate	0.01	0.05
	Iterations	2000	2000
	Subsample	0.8	0.93
	L2_leaf_reg	0.01	0.01
	Rsm	0.81	0.48
	Random_seed	7	19
(6) XGB	Max_depth	5	3
	Eta	0.05	0.05
	N_estimators	1000	1000
	Gamma	0.01	0.01
	Lambda	0.24	0.08
	Subsample	0.5	0.47
	Reg_alpha	0.1	0.04
	Min_child_weight	1	3
	Colsample_bytree	0.6	0.6
(7) ANN	Epochs	10000	10000
	Learning_rate	0.005	0.005
	Hidden layers	2	2
	Number of neurons	32	128
	Activation_function	PReLU	PReLU
	Dropout	0.2	0.2
(8) DNN	N-epoch	5000	5000
	Learning_rate	0.001	0.005
	Hidden layers	3	3
	Kernel_Regularizer_L2	0.01	0.01
	Number of neurons	32	32
	Activation_function	ReLU	ReLU
(9) RF	N_estimators	60	85
	Max_depth	5	10
	Min_samples_split	6	5
	Min_samples_leaf	2	1
	Min_weight_fraction_leaf	0.01	0.01
	Random_state	29	57
(10) SVM	Cache_size	100	10
	Gamma	0.00003	0.0001
	Kernel	Rbf	Rbf
	C	14	10
	Epsilon	0.05	0.07

Table S5 Mean values of performance metrics utilizing five-fold cross-validation

Algorithms	Datasets	Dataset II (Datasets = 316)		Dataset III (Datasets = 813)	
		MSE_{cv}	R^2_{cv}	MSE_{cv}	R^2_{cv}
LGBM-CatB	Training	0.003	0.99	0.001	0.99
	Validation	0.04	0.89	0.02	0.91
	Test	0.08	0.86	0.02	0.93
LGBM-XGB	Training	0.02	0.95	0.001	0.99
	Validation	0.06	0.85	0.03	0.90
	Test	0.06	0.86	0.02	0.93
LGBM	Training	0.001	0.99	0.001	0.99
	Validation	0.04	0.90	0.02	0.92
	Test	0.08	0.86	0.03	0.91
LGBM-RF	Training	0.03	0.94	0.01	0.96
	Validation	0.07	0.84	0.04	0.88
	Test	0.05	0.85	0.02	0.91
CatB	Training	0.02	0.95	0.001	0.99
	Validation	0.07	0.84	0.03	0.91
	Test	0.06	0.84	0.03	0.90
XGB	Training	0.002	0.99	0.002	0.99
	Validation	0.05	0.97	0.03	0.91
	Test	0.08	0.85	0.03	0.90
ANN	Training	0.06	0.86	0.03	0.90
	Validation	0.06	0.86	0.03	0.88
	Test	0.07	0.80	0.03	0.87
DNN	Training	0.03	0.92	0.06	0.79
	Validation	0.10	0.74	0.07	0.78
	Test	0.10	0.75	0.03	0.84
RF	Training	0.05	0.87	0.03	0.90
	Validation	0.09	0.79	0.05	0.83
	Test	0.09	0.75	0.04	0.82
SVM	Training	0.04	0.90	0.02	0.95
	Validation	0.15	0.65	0.06	0.78
	Test	0.13	0.63	0.06	0.75

Table S6 The detailed statistics of the additional instances from the four reported literatures.

Ion	$CeEDTA^-$	$CeEDTA^-$	$CeEDTA^-$	$CeEDTA^-$	ReO_4^-	ReO_4^-	ReO_4^-	ReO_4^-
Clays	Zhisin-bent	Zhisin-bent	Zhisin-bent	Zhisin-bent	MX-80	MX-80	MX-80	MX-80
I (mol/L)	0.67	0.67	0.67	0.67	0.5	0.5	0.5	0.5
ρ_d (kg/m^3)	1300	1600	1700	1800	1300	1500	1600	1700
m (–)	0.368	0.368	0.368	0.368	0.85	0.85	0.85	0.85
K_d (m^3/kg)	1.18×10^{-3}	1.08×10^{-3}	9.7×10^{-4}	8.1×10^{-4}	0	0	0	0
z (–)	–1	–1	–1	–1	–1	–1	–1	–1
pH (–)	5.1	5.1	5.1	5.1	4	4	4	4
T (°C)	20	20	20	20	25	25	25	25
A_{ext} (m^2/g)	35.2	35.2	35.2	35.2	38	38	38	38
ρ_s (kg/m^3)	2670	2670	2670	2670	2760	2760	2760	2760
MW (g/mol)	432.34	432.34	432.34	432.34	250.2	250.2	250.2	250.2
D_w (m^2/s)	5×10^{-10}	5×10^{-10}	5×10^{-10}	5×10^{-10}	1.46×10^{-9}	1.46×10^{-9}	1.46×10^{-9}	1.46×10^{-9}
ε_{tot} (–)	0.513	0.401	0.363	0.326	0.529	0.457	0.420	0.384
α (–)	2.047	2.129	2.012	1.784	0.493	0.415	0.376	0.337
r (Å)	4.59×10^{-10}	4.59×10^{-10}	4.59×10^{-10}	4.59×10^{-10}	1.60×10^{-10}	1.60×10^{-10}	1.60×10^{-10}	1.60×10^{-10}
λ (m^2·S/ mol)	1.91×10^{-3}	1.91×10^{-3}	1.91×10^{-3}	1.91×10^{-3}	5.48×10^{-3}	5.48×10^{-3}	5.48×10^{-3}	5.48×10^{-3}
n_c (–)	32	32	32	32	29	29	29	29
D_e (m^2/s)	1.92×10^{-11}	1.17×10^{-11}	1.09×10^{-11}	8.8×10^{-12}	2.5×10^{-11}	1.9×10^{-11}	1.1×10^{-11}	6.7×10^{-11}
Reference	[2]	[2]	[2]	[2]	[8]	[8]	[8]	[8]

Table S6-Continue.

Ion	ReO_4^-	$HSeO_3^-$	$HSeO_3^-$	$HSeO_3^-$	$HSeO_3^-$	I^-	I^-	I^-
Clays	MX-80	MX-80	MX-80	MX-80	MX-80	Ba-bent	Ba-bent	Ba-bent
I (mol/L)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
ρ_d (kg/m^3)	1800	1500	1600	1700	1800	1300	1400	1500
m (–)	0.85	0.85	0.85	0.85	0.85	0.78	0.78	0.78
K_d (m^3/kg)	0	5.5×10^{-4}	4.1×10^{-4}	3.8×10^{-4}	3.8×10^{-4}	0	0	0
z (–)	–1	–1	–1	–1	–1	–1	–1	–1
pH (–)	4	4	4	4	6	3.13	3.13	3.13
T (°C)	25	25	25	25	25	15	15	15
A_{ext} (m^2/g)	38	38	38	38	38	27.3	27.3	27.3
ρ_s (kg/m^3)	2760	2760	2760	2760	2760	2710	2710	2710
MW (g/mol)	250.2	127.97	127.97	127.97	127.97	126.9	126.9	126.9
D_w (m^2/s)	1.46×10^{-9}	1.55×10^{-9}	1.55×10^{-9}	1.55×10^{-9}	1.55×10^{-9}	2×10^{-9}	2×10^{-9}	2×10^{-9}
ε_{tot} (–)	0.348	0.457	0.420	0.384	0.348	0.520	0.483	0.446
α (–)	0.298	1.282	1.076	1.030	1.032	0.493	0.454	0.415
r (Å)	1.60×10^{-10}	1.51×10^{-10}	1.51×10^{-10}	1.51×10^{-10}	1.51×10^{-10}	1.13×10^{-10}	1.13×10^{-10}	1.13×10^{-10}
λ (m^2·S/ mol)	5.48×10^{-3}	5.82×10^{-3}	5.82×10^{-3}	5.82×10^{-3}	5.82×10^{-3}	7.77×10^{-3}	7.77×10^{-3}	7.77×10^{-3}
n_c (–)	29	29	29	29	29	48	48	48
D_e (m^2/s)	5.1×10^{-11}	1.4×10^{-11}	9.5×10^{-12}	6.4×10^{-12}	4.9×10^{-12}	1.17×10^{-10}	1.03×10^{-10}	8.5×10^{-11}
Reference	[8]	[8]	[8]	[8]	[8]	[9]	[9]	[9]

Table S6-Continue.

Ion	I^-	I^-	$HCrO_4^-$	$HCrO_4^-$	$HCrO_4^-$	$HCrO_4^-$	$HCrO_4^-$	$HCrO_4^-$
Clays	Ba-bent	Ba-bent	Anji	Anji	Anji	Anji	Anji	Anji
I (mol/L)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
ρ_d (kg/m^3)	1600	1700	1200	1300	1400	1500	1600	1700
m (–)	0.78	0.78	0.524	0.524	0.524	0.524	0.524	0.524
K_d (m^3/kg)	0	0	0	0	0	0	0	0
z (–)	–1	–1	–1	–1	–1	–1	–1	–1
pH (–)	3.13	3.13	7.5	7.5	7.5	7.5	7.5	7.5
T (°C)	15	15	20	20	20	20	20	20
A_{ext} (m^2/g)	27.3	27.3	60.3	60.3	60.3	60.3	60.3	60.3
ρ_s (kg/m^3)	2710	2710	2800	2800	2800	2800	2800	2800
MW (g/mol)	126.9	126.9	117.01	117.01	117.01	117.01	117.01	117.01
D_w (m^2/s)	2×10^{-9}	2×10^{-9}	1.13×10^{-9}	1.13×10^{-9}	1.13×10^{-9}	1.13×10^{-9}	1.13×10^{-9}	1.13×10^{-9}
ε_{tot} (–)	0.410	0.373	0.571	0.536	0.500	0.464	0.429	0.393
α (–)	0.376	0.337	0.532	0.493	0.454	0.415	0.376	0.337
r (Å)	1.13×10^{-10}	1.13×10^{-10}	2.03×10^{-10}	2.03×10^{-10}	2.03×10^{-10}	2.03×10^{-10}	2.03×10^{-10}	2.03×10^{-10}
λ (m^2·S/ mol)	7.77×10^{-3}	7.77×10^{-3}	4.32×10^{-3}	4.32×10^{-3}	4.32×10^{-3}	4.32×10^{-3}	4.32×10^{-3}	4.32×10^{-3}
n_c (–)	48	48	16	16	16	16	16	16
D_e (m^2/s)	6.18×10^{-11}	4.2×10^{-11}	8×10^{-11}	6.4×10^{-11}	4.4×10^{-11}	2.8×10^{-11}	1.8×10^{-11}	1.1×10^{-11}
Reference	[9]	[9]	[9]	[9]	[9]	[9]	[9]	[9]

Table S6-Continue.

Ion	$HCrO_4^-$	$CoEDTA^{2-}$	$CoEDTA^{2-}$	$CoEDTA^{2-}$	$CoEDTA^{2-}$	$CoEDTA^{2-}$	$HSeO_3^-$	$HSeO_3^-$
Clays	Anji	GMZ	GMZ	GMZ	GMZ	GMZ	Ba-bent	Ba-bent
I (mol/L)	0.5	0.2	0.2	0.2	0.2	0.2	0.5	0.5
ρ_d (kg/m^3)	1800	1200	1300	1500	1600	1700	1300	1400
m (–)	0.524	0.754	0.754	0.754	0.754	0.754	0.78	0.78
K_d (m^3/kg)	0	0	0	0	0	0	3.8×10^{-4}	3.5×10^{-4}
z (–)	–1	–2	–2	–2	–2	–2	–1	–1
pH (–)	7.5	6.8	6.8	6.8	6.8	6.8	3.1	3.1
T (°C)	20	20	20	20	20	20	22	22
A_{ext} (m^2/g)	60.3	25.6	25.6	25.6	25.6	25.6	27.3	27.3
ρ_s (kg/m^3)	2800	2660	2660	2660	2660	2660	2710	2710
MW (g/mol)	117.01	351.17	351.17	351.17	351.17	351.17	127.97	127.97
D_w (m^2/s)	1.13×10^{-9}	5.00×10^{-10}	5.00×10^{-10}	5.00×10^{-10}	5.00×10^{-10}	5.00×10^{-10}	1.55×10^{-9}	1.55×10^{-9}
ε_{tot} (–)	0.357	0.549	0.511	0.436	0.398	0.361	0.520	0.483
α (–)	0.298	0.532	0.493	0.415	0.376	0.337	1.014	0.973
r (Å)	2.03×10^{-10}	4.59×10^{-10}	4.59×10^{-10}	4.59×10^{-10}	4.59×10^{-10}	4.59×10^{-10}	1.49×10^{-10}	1.49×10^{-10}
λ (m^2·S/ mol)	4.32×10^{-3}	3.82×10^{-3}	3.82×10^{-3}	3.82×10^{-3}	3.82×10^{-3}	3.82×10^{-3}	5.88×10^{-3}	5.88×10^{-3}
n_c (–)	16	54	54	54	54	54	48	48
D_e (m^2/s)	8.5×10^{-12}	1.85×10^{-11}	1.43×10^{-11}	6.8×10^{-12}	6.4×10^{-12}	5.6×10^{-12}	6.8×10^{-11}	5.7×10^{-11}
Reference	[9]	[9]	[9]	[9]	[9]	[9]	[10]	[10]

Table S6-Continue.

Ion	$HSeO_3^-$	$HSeO_3^-$	$HSeO_3^-$	$HSeO_3^-$	$HSeO_3^-$
Clays	Ba-bent	Ba-bent	Ba-bent	GMZ	GMZ
I (mol/L)	0.5	0.5	0.5	0.1	0.3
ρ_d (kg/m^3)	1500	1600	1700	1300	1300
m (–)	0.78	0.78	0.78	0.754	0.754
K_d (m^3/kg)	3.5×10^{-4}	3.26×10^{-4}	3.2×10^{-4}	3.2×10^{-4}	3.4×10^{-4}
z (–)	–1	–1	–1	–1	–1
pH (–)	3.1	3.1	3.1	5.6	5.6
T (°C)	22	22	22	22	22
A_{ext} (m^2/g)	27.3	27.3	27.3	25.6	25.6
ρ_s (kg/m^3)	2710	2710	2710	2660	2660
MW (g/mol)	127.97	127.97	127.97	127.97	127.97
D_w (m^2/s)	1.55×10^{-9}	1.55×10^{-9}	1.55×10^{-9}	1.55×10^{-9}	1.55×10^{-9}
ε_{tot} (–)	0.446	0.410	0.373	0.511	0.511
α (–)	0.971	0.931	0.917	0.927	0.953
r (Å)	1.49×10^{-10}	1.49×10^{-10}	1.49×10^{-10}	1.49×10^{-10}	1.49×10^{-10}
λ (m^2·S/ mol)	5.88×10^{-3}	5.88×10^{-3}	5.88×10^{-3}	5.88×10^{-3}	5.88×10^{-3}
n_c (–)	48	48	48	54	54
D_e (m^2/s)	4.4×10^{-11}	3×10^{-11}	2×10^{-11}	5.6×10^{-11}	6.6×10^{-11}
Reference	[10]	[10]	[10]	[10]	[10]

References

[1] Y. Fukatsu, K. Yotsuji, T. Ohkubo et al., Diffusion of tritiated water, $^{137}Cs^{+}$, and $^{125}I^{-}$ in compacted Ca-montmorillonite: Experimental and modeling approaches. Appl. Clay Sci. **211**, 106176 (2021). https://doi.org/10.1016/j.clay.2021.106176

[2] T. Wu, Y. Hong, D. Shao et al., Experimental and modeling study of the diffusion path of Ce(III)-EDTA in compacted bentonite. Chem. Geol. **636**, 121639 (2023). https://doi.org/10.1016/j.chemgeo.2023.121639

[3] T. Wu, Y. Yang, Z. Wang et al., Anion diffusion in compacted clays by pore-scale simulation and experiments. Water Resour. Res. **56**, e2019WR027037 (2020). https://doi.org/10.1029/2019wr027037

[4] I.C. Bourg, A.C. Bourg, G. Sposito, Modeling diffusion and adsorption in compacted bentonite: a critical review. J. Contam. Hydrol. **61**, 293–302 (2003). https://doi.org/10.1016/S0169-7722(02)00128-6

[5] P. Vanysek, Ionic conductivity and diffusion at infinite dilution. In: CRC Handbook of Chemistry and Physics. **83**, 76–78 (2000).

[6] Y. Ishii, T. Murakami, N. Ohtori, Molecular size and shape effects: Tracer diffusion and the Stokes-Einstein relation. J. Mol. Liq. **346**, 118235 (2022). https://doi.org/10.1016/j.molliq.2021.118235

[7] C. Tournassat, C. Appelo, Modelling approaches for anion-exclusion in compacted Na-bentonite. Geochim. Cosmochim. Acta. **75**, 3698–3710 (2011). https://doi.org/10.1016/j.gca.2011.04.001

[8] Z. Feng, J. Tian, X. Shi et al., Analyzing porosity of compacted bentonite via through diffusion method. J. Radioanal. Nucl. Chem. **333**, 1185–1193 (2024). https://doi.org/10.1007/s10967-024-09368-y

[9] T. Wu, J. Tian, X. Shi et al., Predicting anion diffusion in bentonite using hybrid machine learning model and correlation of physical quantities. Sci. Total Environ. **946**, 174363 (2024). https://doi.org/10.1016/j.scitotenv.2024.174363

[10] X. Shi, J. Tian, J. Shen et al., Application of machine learning in predicting the apparent diffusion coefficient of Se(IV) in compacted bentonite. J. Radioanal. Nucl. Chem. **333**, 5811–5821 (2024). https://doi.org/10.1007/s10967-024-09637-w